

Analyzing 2 Categorical Variables

- You can create categorical variables... *grouping quant.*
- Categorical Variables are often shown ... *2-way table*

	Frosh	Soph	Jr	Senior	Total
Male					
Female					
Total					<i>n</i>

Identify:

- row variable *gender*
- column variable *grade*
- values of the variables
- total (n)
- cells *- where data is (8)*
- totals

rows: 2
columns: 4
2x4

gender: M or F
grade: Fr, So, Jr, Sr

Example:

	Hospital A	Hostpital B	
Died	63	16	79
Survived	2037	784	2821
	2100	800	2900

- ① % of all ppl. A ? $\frac{2100}{2900}$
- ② % of ppl. that went to A died ? $\frac{63}{2100}$
- ③ out of everyone, what % went to B and surv.? $\frac{784}{2900}$
- ④ of survived, what % B ? $\frac{784}{2821}$

2 types of Distributions

1) Marginal Distributions

- How to make: $\text{margins (totals)} \div N$
- Looking for: overall % of each value of var.

~~ALWAYS~~ in %

- Hospital Example:

- Marginal Distribution for Survival Variable:

$$\text{Died: } 79 / 2900 = 2.72\%$$

$$\text{Surv: } 2821 / 2900 = 97.28\%$$

Example:

	Hospital A	Hospital B	n totals
Died	63	16	79
Survived	2037	784	2821
S totals	2100	800	n 2900

○ **Marginal Distribution for Hospital/Variable:**

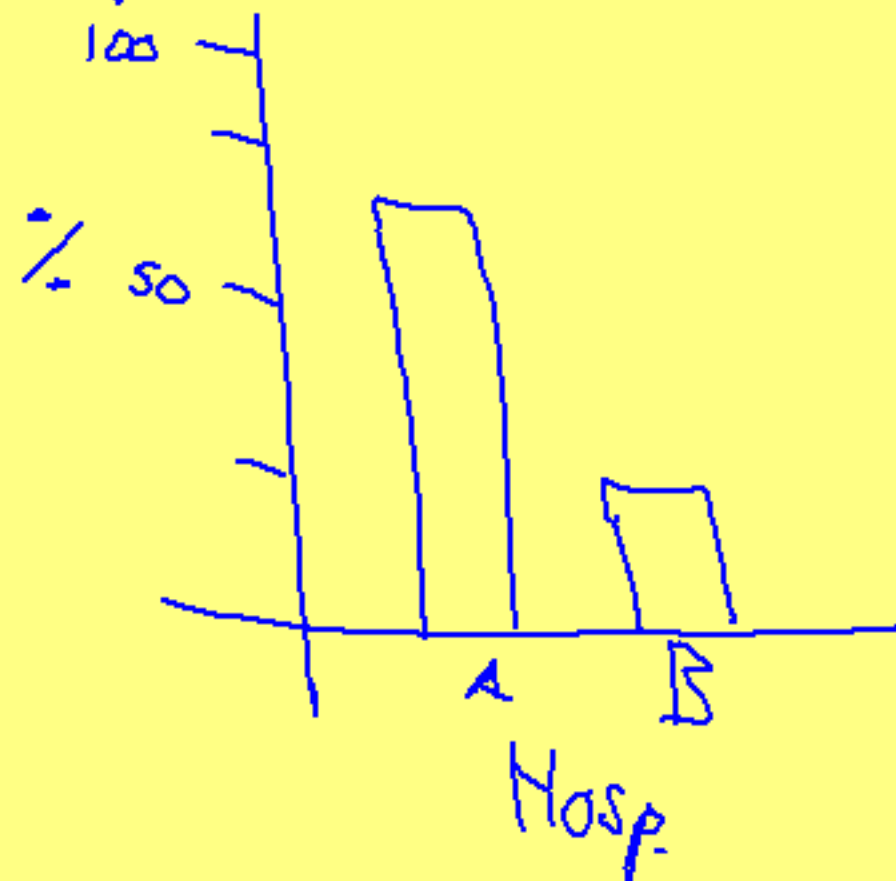
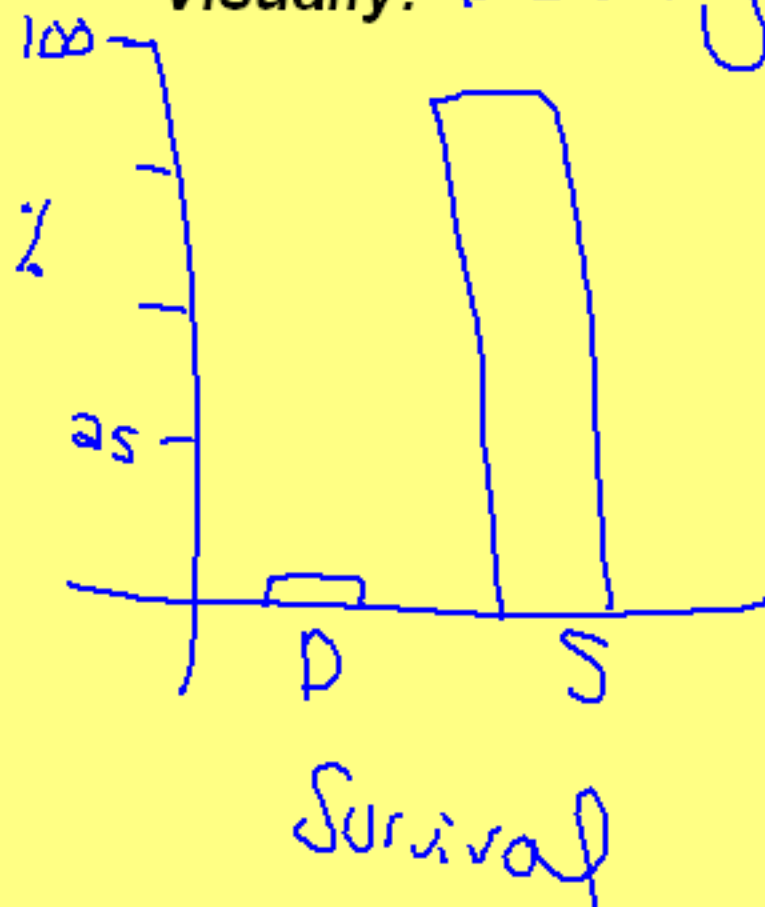
$$A: 2100/2900 = 72.4\%$$

$$B: 800/2900 = 27.6\%$$

Example:

	Hospital A	Hospital B	
Died	63	16	79
Survived	2037	784	2821
	2100	800	2900

Visually: Bar graph



1) Conditional Distributions

- Look at ... one variable
- Then look at ... one value @ a time
- Break down ... value into its parts

* ALWAYS in %

Example:

	Hospital A	Hospital B	
Died	63	16	79
Survived	2037	784	2821
	2100	800	2900

- Hospital Example:

- Conditional Distribution for Survival Variable:

Died

$$A: \frac{63}{79} = 79.7\%$$

$$B: \frac{16}{79} = \frac{20.3\%}{100\%}$$

Survived

$$A: \frac{2037}{2821} = 72.2\%$$

$$B: \frac{784}{2821} = \frac{27.8\%}{100\%}$$

Conditional Distribution for Hospital Variable:

Example:

	Hospital A	Hospital B	
Died	63	16	79
Survived	2037	784	2821
	2100	800	2900

A
D: $63/2100 = 3\%$

S: $2037/2100 = 97\%$
100%

B
D: $18/800 = 2\%$

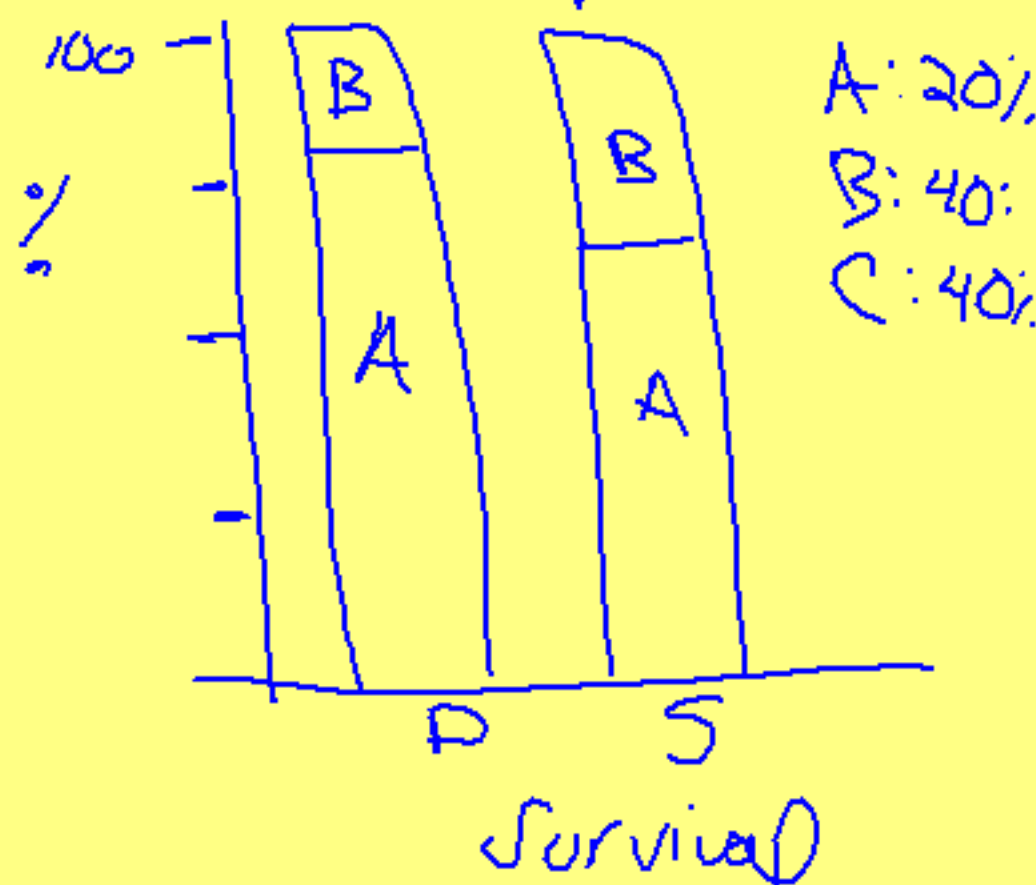
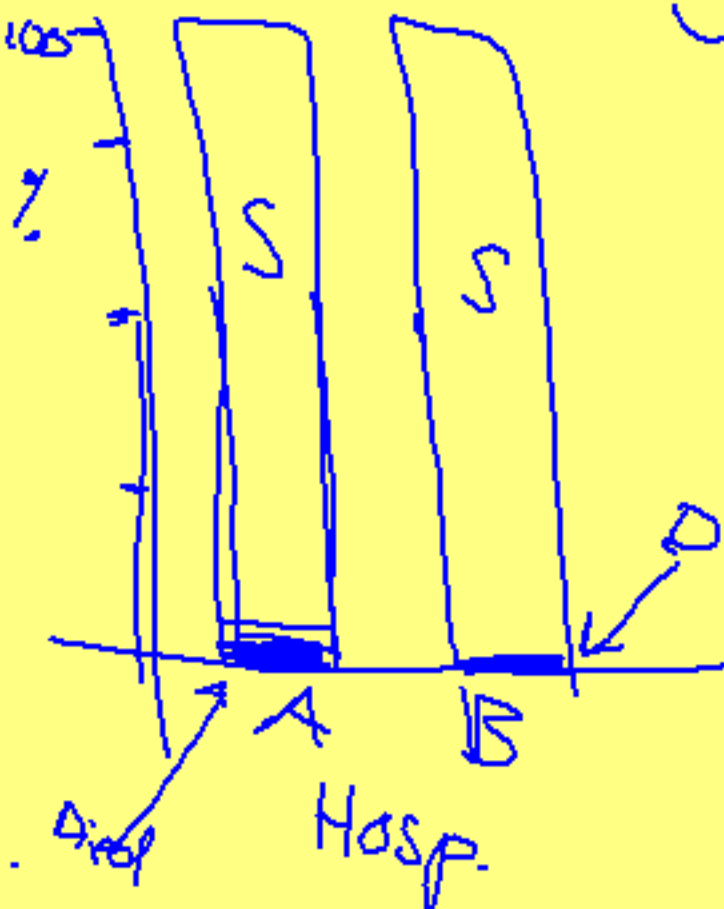
S: $784/800 = 98\%$
100%

Represented Visually:

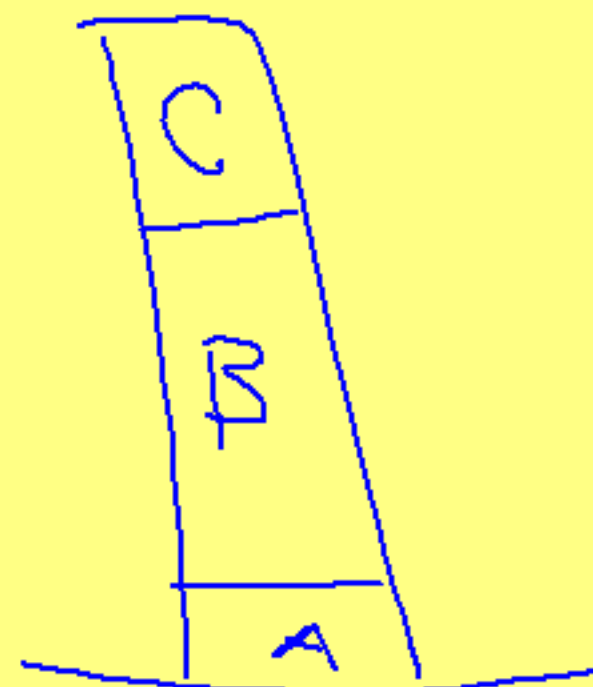
Each bar = 100%

Conditional Variable on ... X-axis

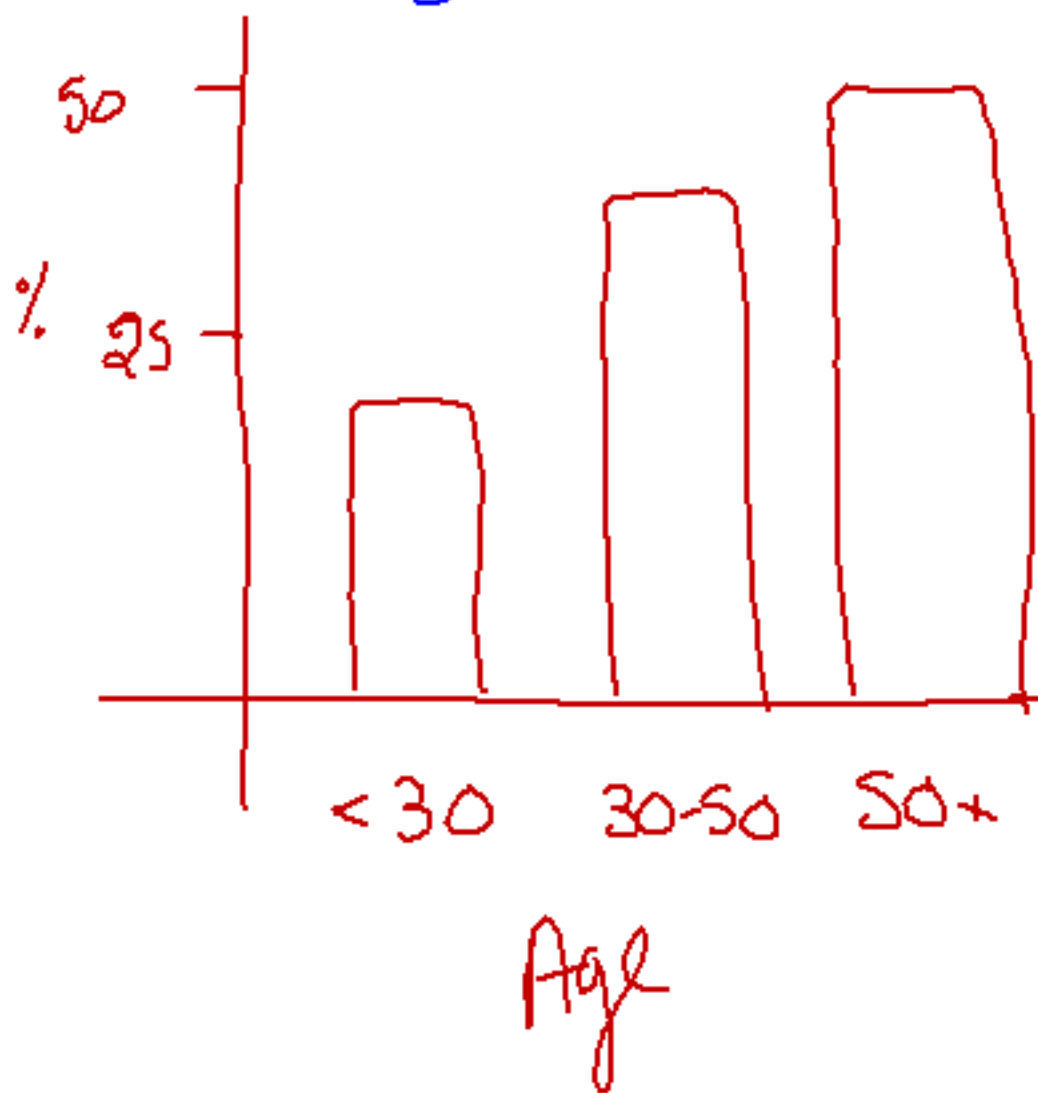
Bars are ... segmented into parts of value



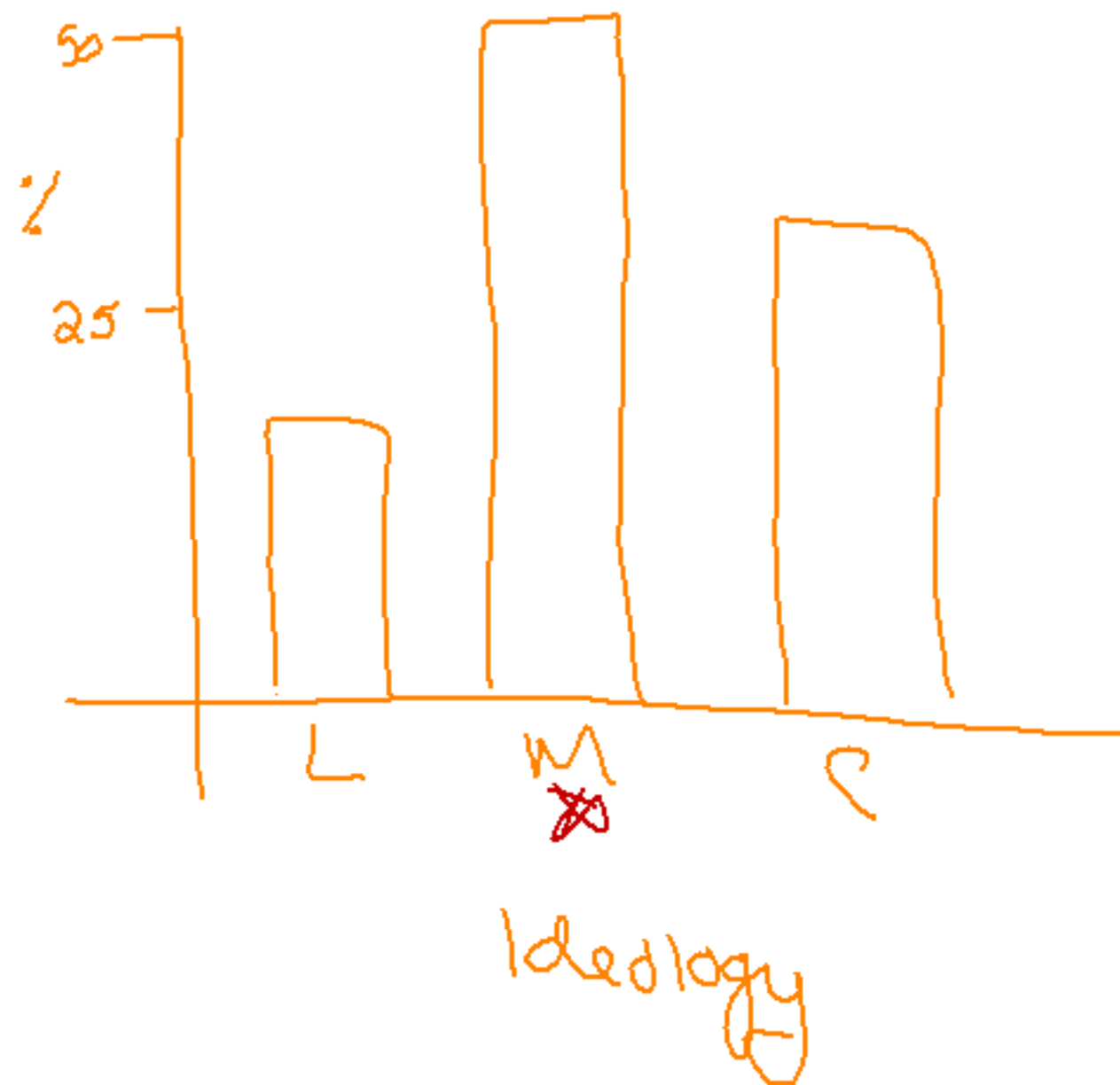
A: 20%
B: 40%
C: 40%



Marginal Age

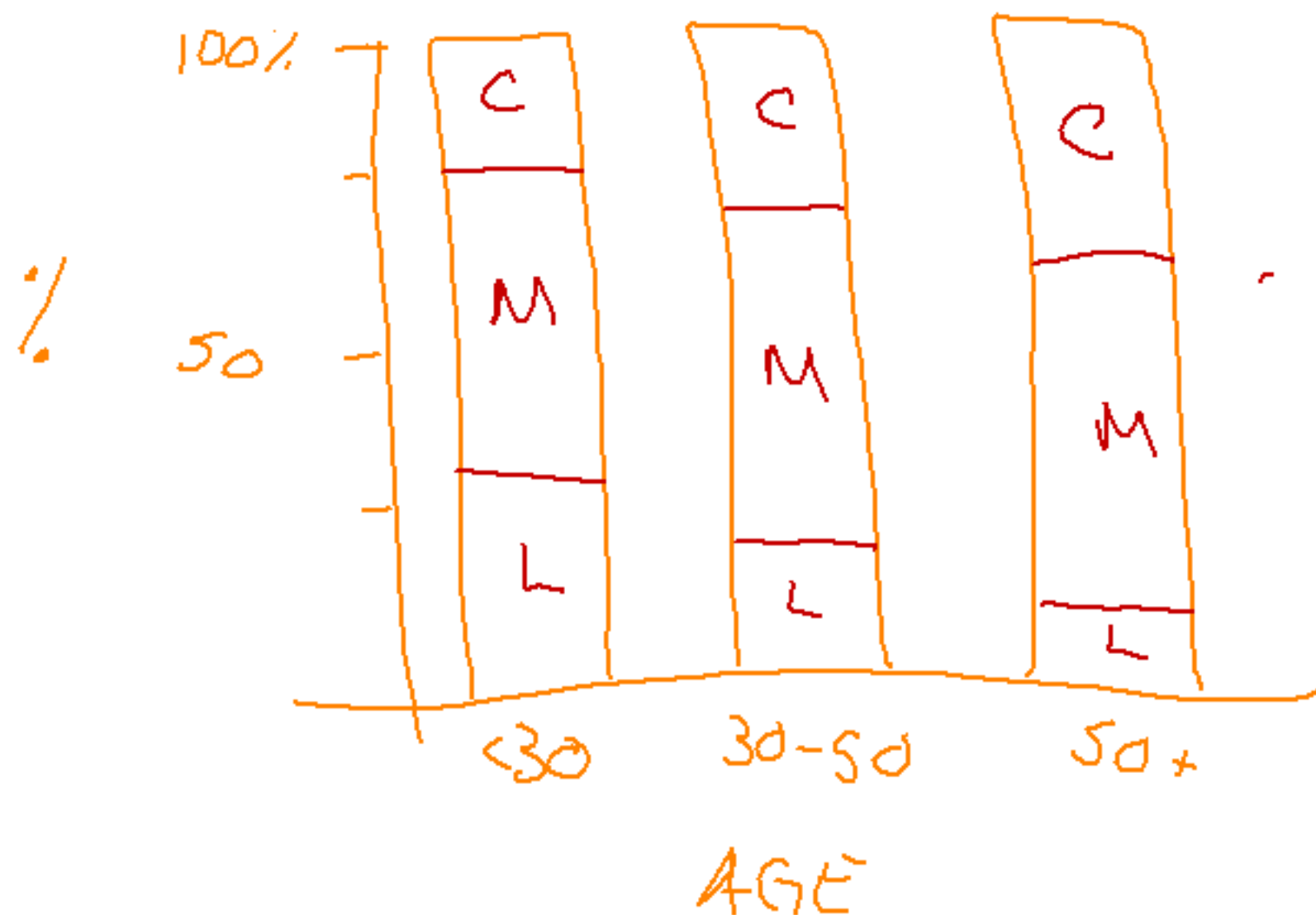


Political Pref.



Conditional - Age

	<u><30</u>	<u>30-50</u>	<u>50+</u>
L	28%	21.3%	15%
M	47.3%	50%	48.5%
C	24.7%	28.8%	36.5%



Independence

Simpson's paradox